International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 25s,2025
https://theaspd.com/index.php

Advanced Image Preprocessing Strategies For Mitigating Challenges In Optical Character Recognition Of Diverse Textual Images

[1]Hidear Talirongan, [2]Bridgett A. Mabuyo, [3]Christine S. Simbajon, [4]Roseclaremath A. Caroro, [5]Markdy Y. Orong, [6]Jerry M. Lumasag, [7]Florence Jean B. Talirongan
[1,2,3,4,5,6,7] College of Computer Studies, Misamis University, Ozamiz City, Philippines, 7200
Email: hidear@mu.edu.ph[1], bridgett.mabuyo@mu.edu.ph[2], christine.simbajon@mu.edu.ph[3], roseclaremath.caroro@mu.edu.ph[4], markdy.orong@mu.edu.ph[5], jerry.lumasag@mu.edu.ph[6], badilles.fj@gmail.com[7]

Abstract
Optical Character Recognition (OCR) is an innovative technology that transforms text in images into text that machines can read. This process is essential for digitizing documents, streamlining data entry, and facilitating text searches within scanned files. However, OCR accuracy is often hindered by poor image quality, noise, and varying text formats, necessitating advanced preprocessing techniques to optimize text clarity and contrast. This study focuses on enhancing OCR accuracy through improved image preprocessing, employing three key methods: Warp Perspective for accurate alignment and cropping of quadrilateral planes; Gaussian Blurring to reduce noise and enhance the distinction between text and background; and Otsu Binarization to convert images into binary format with optimal thresholding. Results indicate significant improvements, with printed text accuracy increasing from 65.56% to 90.35% and mixed text accuracy rising from 41.15% to 56.31%, while handwritten text showed modest gains from 12.27% to 21.08%, reflecting its inherent complexity. These findings demonstrate the potential of advanced preprocessing to enhance OCR performance, particularly for printed and mixed text, while highlighting the need for specialized techniques or models for handwriting recognition. Future research could explore automated cropping using contour detection to improve efficiency and adaptability to diverse document sizes.
Keywords: *EasyOCR; Gaussian Blurring; Otsu Binarization; Text Extraction; Warp Perspective*

1. INTRODUCTION

Optical Character Recognition has become a vital tool for extracting text from images (Shashidhara & Amith, 2024; Fateh et al., 2024; Tang et al., 2024). However, OCR accuracy suffers when the image has poor quality and complex text, like mixed text that contains machine-printed and handwritten characters, as these images frequently have poor quality issues (Singh et al., 2023; Alhamad et al., 2024). To overcome these challenges, advanced image preprocessing techniques are applied to enhance the quality of images, making it easier for OCR to improve the accuracy of recognizing and extracting text (Patience et al., 2024; Ponnuru & Likhitha, 2024; Singh et al., 2025).

Image preprocessing involves enhancing the quality of images before applying OCR. This step improves OCR accuracy by making text more transparent and distinct from the background (Maliński & Okarma, 2023; Kumar et al., 2024). Recent studies in image preprocessing have focused on developing advanced algorithms to address challenges in OCR, particularly for images with mixed or degraded text. Preprocessing methods such as binarization, histogram equalization, and noise reduction have significantly improved OCR performance by creating more precise and uniform text images. These techniques aim to optimize the visibility of textual features, making it easier for OCR algorithms to distinguish characters from the background and other image elements (SOUAHI, 2023).

Studies have shown that advanced preprocessing techniques can reduce error rates in OCR applications by removing unwanted artifacts and enhancing key visual features. This study aims to improve images using advanced preprocessing techniques like Line Warp Perspective, Gaussian Blur, and Otsu Binarization to enhance OCR's accuracy on mixed text images. Binarization is a method of separating a subject into two parts using a threshold value. The Otsu algorithm uses statistical values to determine the thresholds. It is an algorithm used to find gray values in the range of 0 to 255, which minimizes the cost function (Han et al., 2021). By optimizing the preprocessing stage, this research contributes to refining

OCR accuracy in scenarios where conventional techniques fall short, specifically in images with non-uniform text.

The significance of this study lies in its potential to expand the scope of OCR applications across diverse fields, particularly those where accurate text extraction from complex documents is critical, such as in legal, academic, and administrative sectors. By demonstrating the value of

Thresholding methods, this research may provide a foundation for more effective and standardized OCR preprocessing protocols, addressing both structural and noise-related challenges that have traditionally hindered OCR performance on mixed-text images. The anticipated outcome is an enhanced OCR preprocessing framework that leverages advanced image processing techniques, significantly improving OCR accuracy in handling complex, mixed-text images.

Advanced preprocessing methods not only clean up the image but also help the OCR engine focus on the actual letters and numbers (Mursari & Wibowo, 2021). In this study, collecting variety of mixed-text images that represent real-world challenges such as scanned forms with both printed and handwritten notes (Mahadevkar et al., 2024). Each image will go through a series of preprocessing steps: correcting the perspective so text lines appear straight, smoothing uneven regions with Gaussian blur, and converting the picture to clear black-and-white using Otsu's thresholding. By applying these steps in a step-by-step pipeline, we expect the text regions to stand out more clearly against their backgrounds, making it easier for OCR software to recognize each character correctly.

The effectiveness of the preprocessing pipeline will be evaluated by comparing OCR results obtained before and after its application (Singh et al., 2025). Accuracy metrics will be based on the proportion of words and characters correctly recognized in each test image (Hicks et al., 2022). This analysis will identify the specific preprocessing steps that contribute most to improved text extraction. A critical assessment of the methods' strengths and weaknesses will follow, along with recommendations for enhancing the workflow in future studies (Liu et al, 2021). Ultimately, the aim is to establish a clear, repeatable framework that enables consistent improvements in OCR performance on challenging mixed-text images.

2. METHODS

The OCR workflow begins with an input dataset, typically consisting of images that contain text to be digitized. The first step, Image Preprocessing, enhances the image quality to ensure optimal results in later stages. Once preprocessed, the image undergoes Image Segmentation, where the visual data is divided into meaningful sections, like lines or words, which can be analyzed individually. The segmented parts are then fed into the Image & Text Processing using the OCR stage, where optical character recognition algorithms interpret the text within the images. Finally, this workflow extracts the text.

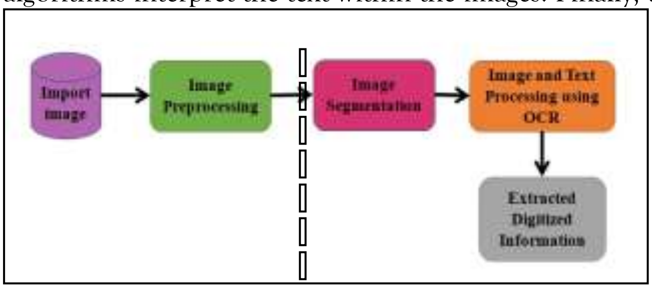

Figure 1. OCR Workflow (Patil et al., 2022)

2.1 Image Preprocessing

2.1.1 Warp Perspective Transform

Warp perspective transform is a geometric transformation technique that adjusts an image's perspective by mapping its points from one plane to another (Meng et al., 2020). This technique is beneficial in OCR preprocessing, as it can straighten or align documents, ensuring a consistent viewpoint for better text extraction. The transformation is defined by a 3x3 matrix, enabling precise control over the image's shape and orientation.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 25s,2025
https://theaspd.com/index.php

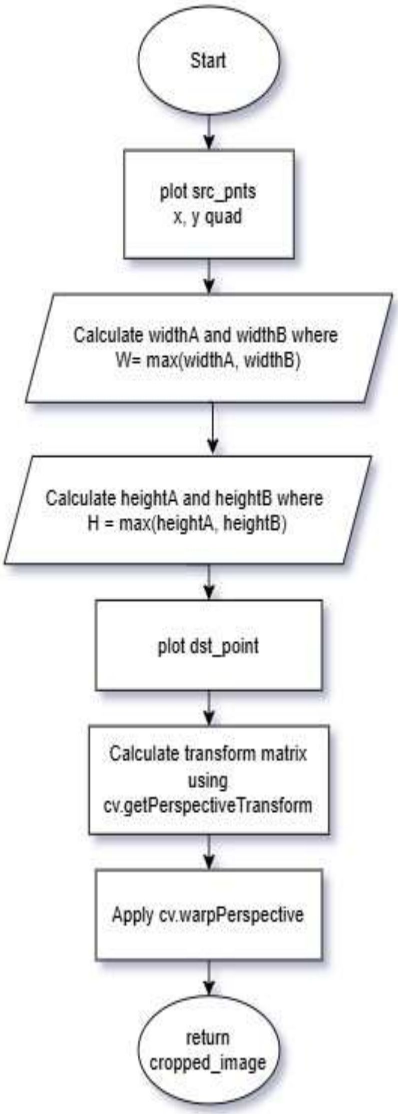

Figure 2. Warp Perspective Transformation flowchart

1. Calculate the source points:

$$src\_points = \begin{bmatrix} x0 & y0 \\ x1 & y1 \\ x2 & y2 \\ x3 & y3 \end{bmatrix}$$

Each row corresponds to a point in the source image, with (x0, y0), (x1, y1), (x2, y2), and (x3, y3) representing the coordinates of the four corners of the region to be transformed.

2. Calculate the width and height of the area:

- $Width(W)$:

$$\text{width}A = \sqrt{(x1 - x0)^2 + (y1 - y0)^2}$$

$$widthB = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}$$

$$W = \max(widthA, widthB)$$

Set W as the maximum of widthA and widthB to ensure that the final width accommodates the longest edge.

- $Height(H)$:

$$\text{heightA} = \sqrt{(x3 - x0)^2 + (y3 - y0)^2}$$

$$heightB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$H = \max(heightA, height)$$

Calculate heightA as the distance between the first and last points and heightB as the distance between the second and third points.

3. Destination points:

$$dst\_points = \begin{bmatrix} (0, & 0) \\ (W - 0, & 0) \\ (W - 1, & H - 1) \\ (0, & H - 1) \end{bmatrix}$$

Specify the destination points in the transformed (output) image, represented as dst_points. These points create a rectangle of width W and height H, aligned with the x- and y-axes.

4. Calculate the perspective transform matrix (M):

$$M = cv.\, getPerspectiveTransform(src\_points, \quad dst\_points)$$

M is a 3x3 matrix that transforms each points (x, y) in src_points to a new points (x′,y′) in dst_points

5. Apply the Perspective Warp to crop the image:

$$cropped\_image = cv.\, warpPerspective(image, \quad M, \quad (W, \quad H))$$

this will output the cropped_image, which is the warped and transformed region of the input image, now aligned to fit within a rectangle of width W and height H.

2.1.2 Gaussian Blur

Gaussian blur is a smoothing technique used in image processing to reduce noise and detail by convolving the image with a Gaussian function. This results in a softened image, where the degree of blurring is controlled by the kernel size and the Gaussian distribution's standard deviation (sigma) (Desai et al., 2020).

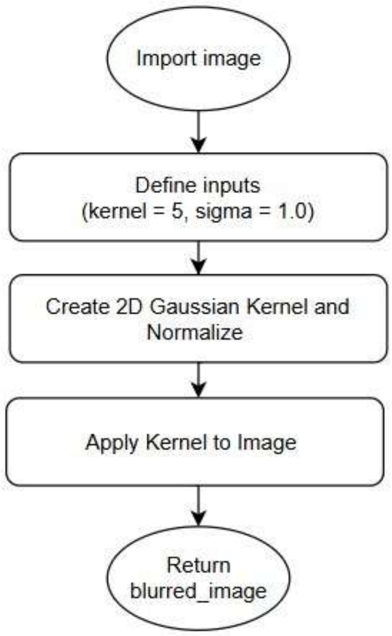

Figure 3. Gaussian Blur Flowchart

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

G(x, y) is the value of the Gaussian function at point (x, y), which becomes the weight. σ is the standard deviation of the Gaussian distribution, which controls the spread or blurriness. x, y coordinates of a point relative to the center of the kernel.

$$e^{-\frac{x^2+y^2}{2\sigma^2}}$$

The exponential term ensures that the weight decreases as you move farther from the center of the kernel.

$$\frac{1}{2\pi\sigma^2}$$

This is the normalization factor, which ensures that the sum of all weights in the kernel is equal to 1.

2.1.3 Otsu Thresholding/Binarization

Otsu binarization is an image thresholding technique that automatically determines the optimal threshold value to separate foreground and background pixels based on their intensity histograms (Cao et al., 2021). By minimizing intra-class variance, Otsu's method ensures effective segmentation, making it widely used in preprocessing for tasks like optical character recognition (OCR) and object detection. It is especially effective for images with bimodal intensity distributions.

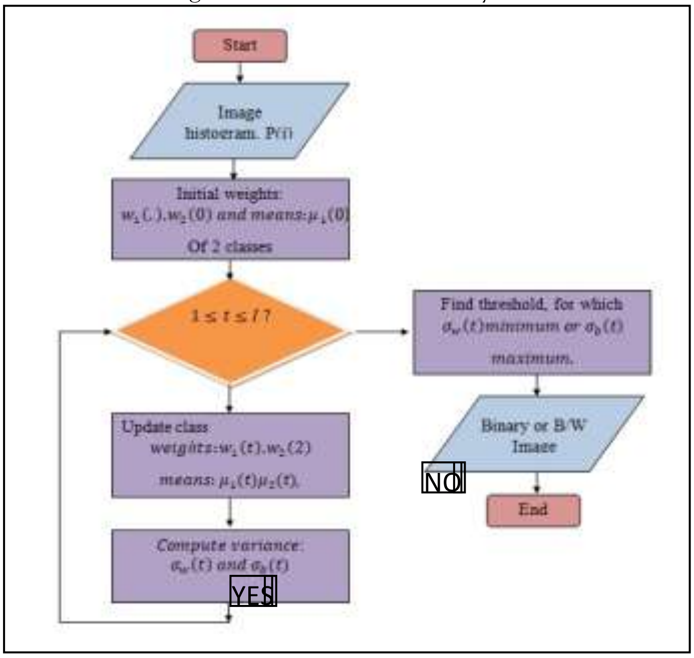

Figure 4. Otsu Binarization Flowchart

2.2 Algorithm Implementation

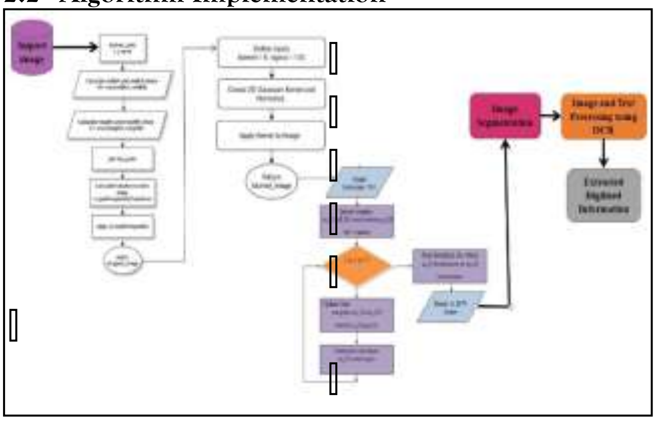

Figure 5. OCR Workflow with Enhanced Image Preprocessing

Figure 5 workflow emphasizes the Image Preprocessing stage as a crucial part of enhancing image quality for optimal OCR accuracy. The process starts with an input dataset, which consists of images that need to be digitized. The Image Preprocessing stage involves three main techniques.

1. Warp Perspective is a method in which you can draw four-sided shapes called quads to define planes in your image and crop them based on the four-point plot.
2. Gaussian Blurring: This method applies a smoothing effect to the image, reducing noise that could interfere with OCR accuracy. By softening the image, Gaussian blurring helps eliminate minor irregularities or artifacts, creating a cleaner, more uniform background and enhancing contrast between text and background.
3. Otsu Binarization: This thresholding technique converts the image into a binary format, highlighting text as distinct black pixels against a white background. Otsu's method automatically finds the optimal threshold to separate foreground and background.

After preprocessing, the image proceeds to Image Segmentation, which is divided into manageable segments, typically lines or words. These segments are then analyzed in the Image & Text Processing using the OCR stage, extracting digitized text, which is then output as Extracted Digitized Information through the EasyOCR library. EasyOCR is a free library in Python that can be used to read text from images. It works with many languages and supports both printed and handwritten text. EasyOCR is easy to use and does not need much setup to extract text (Pattanayak et al., 2023). This preprocessing shows enhancement of OCR performance, ensuring the extracted text is as accurate as possible.

2.3 Evaluation Metrics

2.3.1 Accuracy Score using Lavenshtein Distance

Word Error Rate WER evaluates the accuracy of an OCR system at the word level by measuring the proportion of incorrectly recognized words relative to the reference text; it is computed by dividing the total number of word errors by the total number of words in the reference text (Najam & Faizullah, 2023). WER relies on Levenshtein distance, which calculates the minimum number of edits (insertions, deletions, substitutions) (Zhang et al., 2023).

$$WER = \left(\frac{Lavenshtein\ distance}{total\ number\ of\ words\ in\ the\ reference}\right) 100$$

3. RESULTS AND DISCUSSIONS

3.1 Sample Preprocessed Images

Figures 6, 7, and 8 show the image preprocessing effect using the three techniques. This helps the OCR extract text from images. Figure 8 shows the graph representation of the image histogram and threshold.

A. Machine Printed Text Images

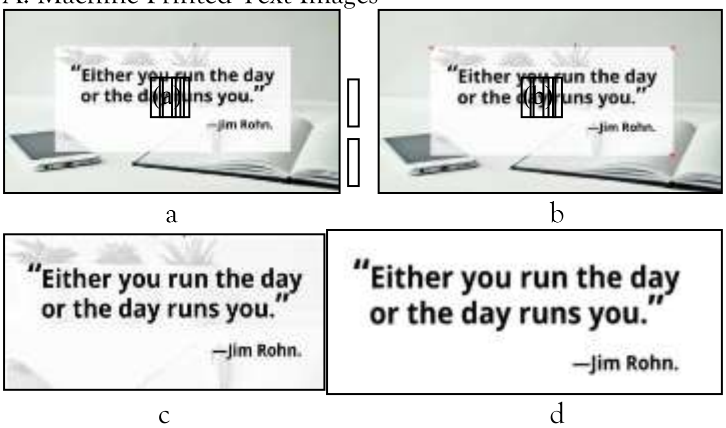

Figure 6. Image Enhancement: (a) raw image; (b) plotted region; (c) cropped image; (d) binarized image

B. Handwritten Text Image

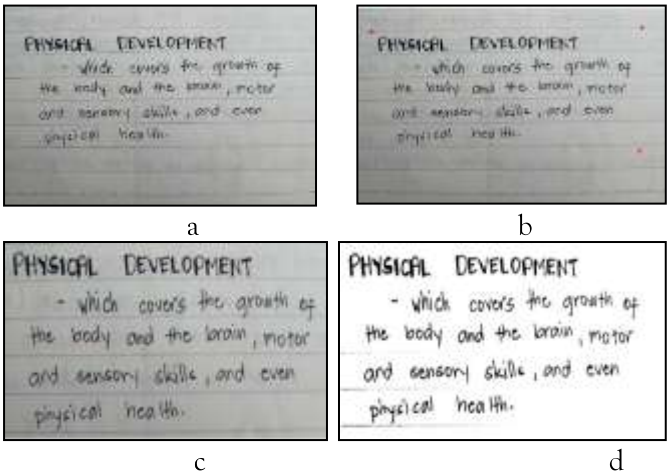

Figure 7. Image Enhancement: (a) raw image; (b) plotted region; (c) cropped image; (d) binarized image

C. Mixed Text Image

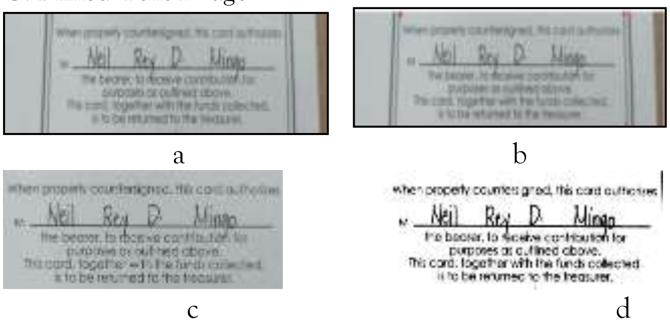

Figure 8. Image Enhancement: (a) raw image; (b) plotted region; (c) cropped image; (d) binarized image

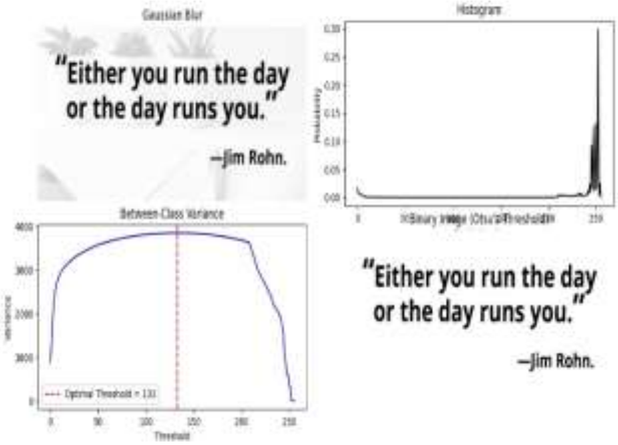

Figure 9. Image histogram and threshold

3.2 Accuracy Test Results

The OCR system's ability to recognize printed text was high in both scenarios. However, enhanced image preprocessing increased the average accuracy significantly from 65.56% to 90.35%. This improvement highlights the effectiveness of advanced preprocessing techniques in improving the clarity and readability of printed text for OCR systems. Handwritten text proved the most challenging for the OCR system, with an average accuracy of only 12.27% without preprocessing. With preprocessing, the accuracy increased to 21.08%, suggesting that preprocessing techniques aid in distinguishing handwriting characteristics. However, the improvement is relatively small, indicating that further adjustments or specialized algorithms may be necessary for effective handwritten text recognition. For mixed text images containing printed and handwritten elements, the OCR system performed at 41.15% accuracy without preprocessing, which increased to 56.31% with enhanced preprocessing. This improvement underscores the potential of advanced image preprocessing to aid OCR in handling complex images with multiple text types. However, further work could focus on optimizing preprocessing for mixed-text photos, as the accuracy gains, while noticeable, still fall short of ideal recognition rates. Overall, enhanced image preprocessing yielded higher accuracy across all text types, with the most significant increase observed in printed text recognition.

Table 1. Accuracy Evaluation of Original OCR and OCR with enhanced image preprocessing

Sample Image	Accuracy					
	OCR			OCR with enhanced image preprocessing		
	Printed Text	Hand written Text	Mixed Text	Printed Text	Hand written Text	Mixed Text
Image1	100%	0.00%	86.67%	100%	50.00%	78.57%
Image2	66.67%	0.00%	16.67%	88.89%	28.57%	16.67%
Image3	66.67%	0.00%	50.00%	75%	10.00%	37.50%
Image4	100%	0.00%	0.00%	100%	0.00%	71.43%
Image5	100%	0.00%	33.33%	100%	11.11%	50.00%
Image6	0.00%	0.00%	72.73%	87.67%	0.00%	60.61%
Image7	76.92	11.11	71.43%	78.75%	11.11	88.57%
Image8	0.00%	66.67	54.55%	100%	0.00%	72.73%
Image9	62.50%	0.00%	8.70%	87.50%	50.00%	30.43%
Image10	82.86%	50.00%	17.39%	85.71%	50.00%	56.62%
Average Accuracy	65.56%	12.27%	41.15%	**90.35%**	**21.08%**	**56.31%**

The preprocessing techniques, including noise reduction and contrast adjustment, effectively improved image quality, making text more straightforward for the OCR system. However, the results indicate that enhanced preprocessing contributes positively to printed and mixed-text recognition, but OCR systems still struggle with handwritten text, likely due to its variability and complexity. The data also reveal that the order of preprocessing steps matters. For instance, applying noise reduction before contrast adjustment led to clearer text edges in several mixed-text images, improving OCR's ability to separate overlapping strokes. In contrast, reversing this order sometimes left faint artifacts that confused the OCR engine. This suggests that a carefully designed sequence, first removing random noise, then boosting contrast, and finally applying thresholding can yield more consistent gains across different image types.

Another observation is the varying impact of preprocessing on different document qualities. Images with strong shadows or uneven lighting saw the largest jumps in printed-text accuracy highest gain, while those taken at steep angles benefited most from perspective correction. However, images that were both low-resolution and heavily handwritten still lagged despite all enhancements. This point to a possible need for adaptive strategies that first assess image quality and then choose the best combination of filters, rather than applying a fixed pipeline to every image.

4. CONCLUSIONS AND RECOMMENDATIONS

The results demonstrate that advanced image preprocessing techniques can considerably improve OCR accuracy, particularly for printed and mixed text. While mixed-text recognition has a noticeable benefit, handwritten text accuracy remains relatively low. This finding highlights the clear advantage of sophisticated preprocessing in standard and mixed-content scenarios and underscores the ongoing challenge of reliably recognizing handwritten material.

Future research should explore targeted enhancements in preprocessing techniques specifically for handwritten text or investigate specialized OCR models designed for handwriting recognition. Implementing automatic cropping via contour detection for varied document types and image sizes would reduce reliance on manual, user-dependent cropping. Further examination of Gaussian blurring moving away from a fixed kernel size toward an adaptive or dynamic approach could better accommodate the variability inherent in individual handwriting styles without compromising performance on other image types. In light of these findings, it is recommended to develop an adaptive preprocessing controller that analyzes each image's key characteristics such as lighting variance, angle distortion, and handwriting density and then selects the optimal sequence of filters. By tailoring the pipeline to each image's specific flaws, OCR systems can achieve better overall performance, especially in challenging mixed-text scenarios.

5. REFERENCES

- Alhamad, H. A., Shehab, M., Shambour, M. K. Y., Abu-Hashem, M. A., Abuthawabeh, A., Al-Aqrabi, H., ... & Shannaq, F. B. (2024). Handwritten recognition techniques: a comprehensive review. *Symmetry*, *16*(6), 681.
- Cao, Q., Qingge, L., & Yang, P. (2021). [Retracted] Performance Analysis of Otsu-Based Thresholding Algorithms: A Comparative Study. Journal of Sensors, 2021(1), 4896853.
- Desai, B., Kushwaha, U., Jha, S., & NMIMS, M. (2020). Image filtering-techniques algorithms and applications. Applied GIS, 7(11), 970-975.
- Fateh, A., Fateh, M., & Abolghasemi, V. (2024). Enhancing optical character recognition: Efficient techniques for document layout analysis and text line detection. *Engineering Reports*, *6*(9), e12832.
- Han, Y., Tarakey, B. A., Hong, S. J., Kim, S. Y., Kim, E., Lee, C. H., & Kim, G. (2021). Calibration and image processing of aerial thermal image for UAV application in crop water stress estimation. Journal of Sensors, 2021(1), 5537795.
- Kumar, S., Sharma, N. K., Sharma, M., & Agrawal, N. (2024). Text extraction from images using tesseract. *Deep Learning Techniques for Automation and Industrial Applications*, 1-18.
- Liu, Z., Cheng, W., Jim, C. Y., Morakinyo, T. E., Shi, Y., & Ng, E. (2021). Heat mitigation benefits of urban green and blue infrastructures: A systematic review of modeling techniques, validation and scenario simulation in ENVI-met V4. *Building and Environment*, *200*, 107939.
- Mahadevkar, Supriya V., Shruti Patil, Ketan Kotecha, Lim Way Soong, and Tanupriya Choudhury. "Exploring AI-driven approaches for unstructured document analysis and future horizons." *Journal of Big Data* 11, no. 1 (2024): 92.
- Maliński, K., & Okarma, K. (2023). Analysis of image preprocessing and binarization methods for OCR-based detection and classification of electronic integrated circuit labeling. Electronics, 12(11), 2449.
- Meng, D., Li, L., Wang, S., Gao, X., Zha, Z. J., & Huang, Q. (2020, October). Finegrained feature alignment with part perspective transformation for vehicle reid. In Proceedings of the 28th ACM International Conference on Multimedia (pp. 619-627).
- Mursari, L. R., & Wibowo, A. (2021). The effectiveness of image preprocessing on digital handwritten scripts recognition with the implementation of OCR Tesseract. *Computer Engineering and Applications Journal*, *10*(3), 177-186.
- Najam, R., & Faizullah, S. (2023). Analysis of recent deep learning techniques for Arabic handwritten-text OCR and post-OCR correction. Applied Sciences, 13(13), 7568.
- Patience, O. O., Amaechi, E. M., George, O., & Isaac, O. N. (2024). Enhanced Text Recognition in Images Using Tesseract OCR within the Laravel Framework. Asian Journal of Research in Computer Science, 17(9), 58-69.
- Patil, S., Varadarajan, V., Mahadevkar, S., Athawade, R., Maheshwari, L., Kumbhare, S., ... & Kotecha, K. (2022). Enhancing optical character recognition on images with mixed text using semantic segmentation. Journal of Sensor and Actuator Networks, 11(4), 63.
- Pattanayak, B. K., Biswal, A. K., Laha, S. R., Pattnaik, S., Dash, B. B., & Patra, S. S. (2023, October). A Novel Technique for Handwritten Text Recognition Using Easy OCR. In 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS) (pp. 1115-1119). IEEE.
- Ponnuru, M., & Likhitha, A. (2024). Image-Based Extraction of Prescription Information using OCR-Tesseract. *Procedia Computer Science*, *235*, 1077-1086.
- Shashidhara, B., & Amith, G. K. (2024, July). A review on text extraction techniques for degraded historical document images. In 2024 Second International Conference on Advances in Information Technology (ICAIT) (Vol. 1, pp. 1-8). IEEE.
- Singh, S., Garg, N. K., & Kumar, M. (2023). Feature extraction and classification techniques for handwritten Devanagari text recognition: a survey. Multimedia Tools and Applications, 82(1), 747-775.
- Singh, K., Sharma, A., Shahi, A. K., Prateek, P., & Munzarin, S. (2025, February). Efficient image preprocessing pipeline for accurate text extraction using OCR and pattern matching. In *AIP Conference Proceedings* (Vol. 3224, No. 1). AIP Publishing.
- SOUAHI, H. (2023). Analytic study of the preprocessing methods impact on historical document analysis and classification (Doctoral dissertation, Université Echahid Chikh Larbi Tébessi-Tébessa).
- Tang, Q., Lee, Y., & Jung, H. (2024). The Industrial Application of Artificial Intelligence-Based Optical Character Recognition in Modern Manufacturing Innovations. *Sustainability*, *16*(5), 2161.
- Zhang, J., Qian, L., Wang, S., Zhu, Y., Gao, Z., Yu, H., & Li, W. (2023). A Levenshtein distance-based method for word segmentation in corpus augmentation of geoscience texts. Annals of GIS, 29(2), 293-306.